

СТАТИСТИЧЕСКИЙ АНАЛИЗ МЕДИЦИНСКОЙ ИНФОРМАЦИИ. ЧАСТЬ ПЕРВАЯ. ПОДГОТОВКА ДАННЫХ. ОПРЕДЕЛЕНИЕ ОБЪЕМА ВЫБОРКИ

Томашевский А. В.*, Каширин В. А.***, Воронцова Л. Л.***

* Запорожский национальный технический университет

** Запорожская медицинская академия последипломного образования

В статье представлены практические рекомендации по использованию компьютерных технологий для статистической обработки данных экспериментальных и/или клинических исследований. Рассмотрены вопросы подготовка данных исследований и определение объема выборки.

Ключевые слова: статистический анализ медицинской информации, выборка.

Обязательным условием оценки результатов экспериментальных и/или клинических исследований является обработка полученных данных с использованием современных методов статистического анализа. Существующие методические и учебные пособия, касающиеся указанного вопроса, перегружены теоретическим материалом, не дают четких практических рекомендаций по работе с компьютерными программами и, что основное, предполагают применение анализа при завершении, но не планировании и начале проведения работы. Вместе с тем, именно на этих этапах возникают вопросы, решение которых способно кардинально облегчить выполнение поставленных задач.

Цель цикла настоящих статей – представить практические рекомендации по использованию компьютерных технологий для статистической обработки данных экспериментальных и/или клинических исследований.

Подготовка данных для статистической обработки предполагает использование программы MS Excel, в которую, в виде строк и столбцов, вносятся значения показателей клинически здоровых лиц (контрольные значения), значения показателей больных, определенные до начала (исходные) и в процессе проведения исследований (опытные значения).

Особенности компьютерной обработки рассмотрим на примере исследования значений процента зрелых Т-лимфоцитов (розеткообразование с CD3-диагностикумом; норма взрослых $60 \pm 5\%$; референтные значения 50–80%) при комбинированном лечении больных раком гор-тани ($T_3N_0M_0$).

Определение минимального объема выборки.

На начальном этапе работы возникает вопрос о количестве больных (объем выборки), которые должны быть включены в исследования, для получения статистически обоснованных выводов.

Любой показатель рассматривается как случайная величина и характеризуется распределением вероятностей значений. Наиболее распространенным законом распределения является нормальный закон [1, 3], при предположении которого, минимальный объем выборки [2, 3], составляет:

$$n = (z_{\alpha} + z_{\beta})^2 \frac{s_e^2 + s_i^2}{\Delta^2}, \quad (1)$$

где z_{α}, z_{β} – критические значения нормального распределения, соответствующие заданным уровням ошибок 1 и 2 рода, которые определяются по таблицам нормального распределения [1, 3] или с помощью функции НОРМСТОБР;

s_e^2, s_o^2 – дисперсии в сравниваемых выборках;

Δ – требуемая величина различий между средними значениями показателей в сравниваемых группах больных.

Компьютерная обработка первичных данных с помощью программы Excel включает в себя следующие шаги.

1. Для расчета значений необходимо активировать кнопку fx и в появившемся окне «Мастер функций» выбрать (щелчком левой клавиши мышки) категорию «статистические» (рис. 1, а). Далее в подокне «выберите функцию» выбрать функцию НОРМСТОБР, нажать ОК, что приведет к открытию окна функции НОРМСТОБР (рис. 1, б), в которое вводится задаваемое значение вероятности $\alpha/2$. Нажать ОК.

Величина α – вероятность ошибки 1-го рода, состоящая в том, что будет отвергнута гипотеза о неразличимости в исследуемых показателях контрольной и исходной выборки. Так как исследуется двухсторонняя область, то значение

α делится на 2. При $\alpha = 0,05$, вводится значение 0,025, и функция возвращает значение -1,96. Таким образом, величина $Z_\alpha = 1,96$ (берется абсолютная величина). Вычисления Z_β выполняется аналогично (при $\beta = 0,05$, получим $Z_\beta = 1,96$).

Приведенную последовательность действий можно представить схемой:

$f_x \rightarrow$ статистические \rightarrow НОРМСТОБР \rightarrow ОК $\rightarrow \alpha/2 \rightarrow$ ОК

2. Для расчета средних значений исследуемых показателей следует мышкой выделить анализируемые столбцы, затем щелкнуть кнопку $\Sigma \blacktriangledown$ и в открывшемся окне выбрать функцию «Среднее». Искомые значения появятся в ячейках непосредственно под столбцами значений показателей.

3. Для расчета дисперсий – выделить ячейку для записи дисперсий – выделить ячейку, щелкнуть по кнопке f_x , в появившемся окне «Мастер функций» выбрать категорию «статистические», выбрать функцию «ДИСП», нажать ОК и в качестве аргументов указать адреса ячеек с исходными данными, т.е. выделить мышкой столбец анализируемых значений, нажать ОК (рис. 3). Приведенную последовательность действий можно представить схемой:

$f_x \rightarrow$ статистические \rightarrow ДИСП \rightarrow ОК \rightarrow выделить столбец ячеек \rightarrow ОК

4. Величину $\Delta = 20$ – можно оценить как разность полученных средних контрольной и опытной групп больных.

5. Для расчета величины объема выборки в окне формул программы Excel вводится выражение, соответствующее формуле (1). Для рассматриваемого примера расчет показан на рисунке 2. Определено, что $n = 6,9$, т.е. минимальный объем выборки для определения различия между контрольными и исходными значениями показателей равен 7. Данный расчет можно представить схемой:

$f_x \rightarrow$ выделить ячейку для отображения $n \rightarrow$ ввести формулу (1) в окно формул \rightarrow ОК

Исходные данные для расчета:

- в ячейки F4, F6 необходимо ввести в значения Z_α, Z_β , соответственно;
- в ячейки F9 необходимо ввести в значение Δ ;
- из ячеек B34, C34 использовать значения S_e^2, S_o^2 , соответственно.

Результат расчета отображается в ячейке F12, для чего эту ячейку необходимо выделить щелчком мыши, затем щелкнуть мышью по окну формул, после чего в нем появится мигающий маркер. Ввести в окно формул следующее выражение:

$$=(F4+F6)^2*((B34+C34)/F9^2), \quad (2),$$

где знак «=» означает, что в ячейку вводится формула;

«^2» — означает возведение в квадрат;

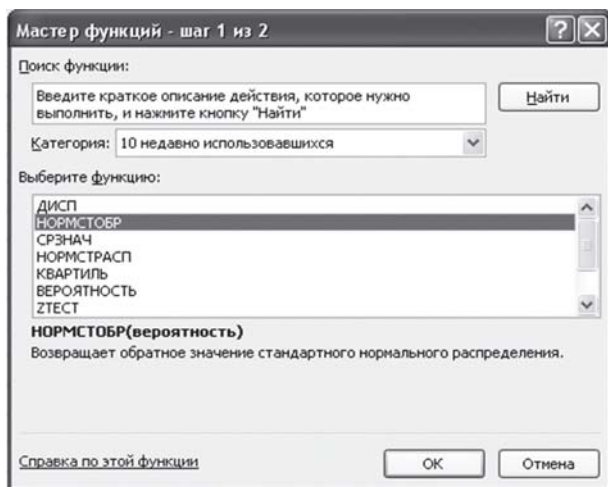
символы «*», «/» — означают умножение и деление.

После ввода нажать клавишу Enter и в ячейке отобразится результат расчета – значение объема выборки n .

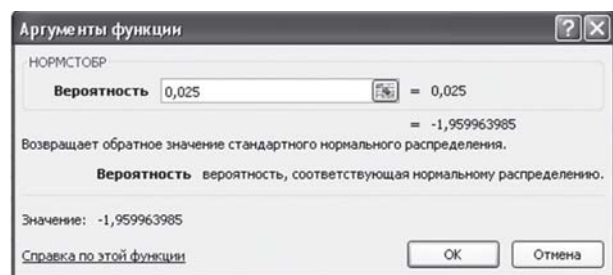
Таким образом, при вводе в ячейки F4, F6, B34, C34, F9 соответствующих исходных значений и в ячейку F12 выражения (2), получим результат расчета по формуле (1).

Аналогичный расчет выполняется по всем остальным показателям (к примеру, CD4+; CD8+ и др.). Из полученных значений выбирается максимальное, которое и будет определять минимальное количество больных, необходимых для данного клинического исследования.

Принято считать, что при $n \geq 60$, выборка большая и репрезентативная, а при $n < 60$ – малая. Однако такое разделение выборок носит достаточно условный характер. Понятие репрезентативности не обязательно связано с объемом выборки, но в большей степени зависит от



а)



б)

Рис. 1. а) Окно «Мастер функций»; б) Окно функции НОРМСТОБР

F12			f ₉ = (F4+F6)^2*(B34+C34)*F9^2				
	A	B	C	D	E	F	G
1	№ пп	Контрол.	Опытн.				
2	1	64	58				
3	2	62	63				
4	3	62	27				
5	4	68	26				
6	5	60	44				
7	6	63	30				
8	7	64	44				
9	8	62	43				
10	9	65	18				
11	10	66	40				
12	11	62	40				
13	12	65	60				
14	13	62	64				
15	14	64	37				
16	15	62	39				
17	16	60	68				
18	17	64	35				
19	18	58	60				
20	19	65	44				
21	20	66	34				
22	21	62	29				
23	22	60	49				
24	23	64	52				
25	24	66	48				
26	25	68	54				
27	26	61	28				
28	27	59	36				
29	28	65	54				
30	29	64	38				
31	30	62	64				
32	31	67	33				
33	Среднее	63,3	43,8				
34	дисперсия	6,413	174,140				

$\sum \alpha =$	1,96
$\sum \beta =$	1,96
$\Delta =$	20
n =	6,9

Рис. 2. Определение объема выборки с помощью программы MS Excel

исследуемого объекта, объема генеральной совокупности, трудоемкости и стоимости проведения исследований. Зачастую генеральная совокупность, в силу тех или иных обстоятельств ограничена.

При этом, выборка может быть еще меньшей. Таким образом, выборка близкая по объему к объему к генеральной совокупности, может считаться малой, но репрезентативной [1].

Следовательно, если для рассмотренного выше примера, количество больных, равное 7, достаточно для формирования статистически обоснованного заключения о различии исследуемого показателя у здоровых и больных пациентов. Можно предположить, что в процессе исследований динамики значений CD3+, количество больных равное 7 достаточно для формирования групп и получения статистически обоснованных выводов.

Заключение

Использование компьютерных технологий требует внесения первичных данных исследований в программу Excel в виде строк и столбцов. Для определения различий между средними значениями сравниваемых групп пациентов достаточным будет объем выборки, рассчитанный по формуле (1). Если предположения, принятые при использовании формулы (1), будут нарушены, в частности закон распределения будет отличен от нормального, исследуются корреляционные или иные зависимости, то для определения объема выборки необходимо использовать другие подходы.

В следующих статьях предполагается подробно рассмотреть практическое использование решения типовых задач с конкретными примерами проверки статистических гипотез.

Литература

1. Томашевський О. В., Рісіков В. П. Комп'ютерні технології статистичної обробки даних. – Запоріжжя: Запорізький національний технічний університет, 2006. – 175 с.
2. Планирование медицинских экспериментов. Ч. 2: Расчет объемов выборок / StatSoft, Inc. (2001). Электронный учебник по статистике. Москва, StatSoft. WEB: <http://www.statsoft.ru/home/textbook/default.htm>
3. Методы обработки медицинской информации / Минцер О. П., Угаров Б. Н., Власов В. В. Киев: Высшая школа. Головное изд-во, 1982.– с. 31.

СТАТИСТИЧНИЙ АНАЛІЗ МЕДИЧНОЇ ІНФОРМАЦІЇ. ЧАСТИНА ПЕРША. ПІДГОТОВКА ДАНИХ. ВИЗНАЧЕННЯ ОБ'ЄМУ ВИБІРКИ

Томашевський О. В. , Каширін В. О.** , Воронцова Л. Л.***

** Запорізький національний технічний університет*

*** Запорізька медична академія післядипломної освіти*

В статті представлені практичні рекомендації з використання комп'ютерних технологій для статистичної обробки даних експериментальних і/або клінічних досліджень. Розглянуті питання підготовки даних досліджень і визначення об'єму вибірки.

Ключові слова: статистичний аналіз медичної інформації, вибірка.

STATISTICAL ANALYSIS OF THE MEDICAL INFORMATION

*Tomashevsky O. V. * , Kashirin V. O.** , Vorontsova L. L. ***

** Zaporizhzhya National Technical University*

*** Zaporizhzhya Medical Academy of Postgraduate Education*

The article practical recommendations are presented on the use of computers technologies for statistical treatment of these experimental and/or clinical researches. In the first part of work questions are considered preparation of these researches and determination of sample size.

Key words: medical information statistical analysis, sample size.